Optimal Estimation of Multicast Membership

Sara Alouf*, Eitan Altman, *senior member*, *IEEE*, Chadi Barakat and Philippe Nain

INRIA Sophia Antipolis, B.P. 93
06902, Sophia Antipolis Cedex, France
{salouf, altman, cbarakat, nain}@sophia.inria.fr

Abstract

This paper addresses optimal on-line estimation of the size of a multicast group. Three distinct approaches are used. The first one builds on Kalman filter theory to derive the MSE-optimal estimator in heavy-traffic regime. Under more general assumptions, the second approach uses Wiener filter theory to compute the MSE-optimal linear filter. The third approach develops the best first-order linear filter from which an estimator that holds for any on-time distribution is derived. Our estimators are tested on real video traces and exhibit good performance. The paper also provides guidelines on how to tune the parameters involved in the schemes in order to achieve high quality estimation while simultaneously avoiding feedback implosion.

keyworkds— 2-ESTM, 2-SDES, Signal Processing in Networking

1 Introduction

Since its introduction, IP multicast [8, 9] has seen slow deployment in the Internet. As stated in [10], the service model and architecture do not efficiently provide or address many features required for a robust implementation of multicast. However, the fact remains that IP multicast is very appealing in offering scalable point-to-multipoint delivery especially in satellite communications. This work is motivated by the conviction that large-scale multicast applications will soon be deployed in the Internet. We believe that membership estimates will be an essential component of this widespread deployment as they can be very useful for scalable multicast. Future Internet radios and TVs will need to characterize their audience preferences and to follow the fluctuations of the audience size over time. Dutta, Schulzrinne and Yemini proposed an architecture for Internet radio and TV called MarconiNet [11] that relies on RTCP [21, 22], the real-time transport control protocol in the Internet. Even though RTCP provides an easy mechanism for collecting statistics on the size of the audience, it does not scale well to large multicast sessions. In such applications, sampling-based techniques are more appropriate.

There has been a significant research effort in devising sampling-based schemes for the estimation of the membership in multicast sessions [5, 12, 17, 19] (see also [2, Ch. 2] where the main features of these schemes are presented). However, none of these schemes have been shown to be optimal within some particular set; further, at the exception of the scheme in [19], they do not use past information, an essential feature in estimation theory.

In this work, we propose a novel sampling-based technique that we now describe. Whenever a source is interested in knowing how many receivers are connected to the multicast session (or are actively following some application that is being broadcasted), it asks all connected members or participants to send an acknowledgment (ACK) every S seconds. However, in order to avoid that too many ACKs are sent to the sources in the case of a large multicast group, a phenomenon refers to as *feedback implosion*, each participant only sends an ACK every S seconds with probability p. Clearly, the values of p and S will have a direct impact on the quality of the estimator and on the number of ACKs that are travelling to the source. Ideally, p should be large and S should be small so that the source collects enough *correlated* observations for its (whatever) estimation scheme to work efficiently. But this ideal scenario would yield feedback implosion. The challenge is therefore to design an estimation scheme for the size of the multicast audience that is accurate without generating too many ACKs.

Throughout the paper, we address the issue of estimating the membership of a multicast group. We build on adaptive filtering theory to derive the estimator. Three distinct approaches are successively considered, based on Kalman filtering theory, Wiener filtering theory and least square estimation, respectively.

The Kalman filter provides a linear, unbiased, and minimum error variance recursive algorithm to optimally estimate the unknown state of a linear dynamic system from noisy data taken at discrete real-time intervals. Furthermore, under normality assumptions, this filter is optimal, not only among all linear filters based on a set of observations, but among all measurable filters [18, 23]. Since our measurements are collected at discrete times, Kalman filter therefore appears as an appealing approach for solving our estimation problem. In Section 4 we show that under some conditions (heavy traffic regime and exponential on-times – the on-time is defined as the length of time during which a user participates to a multicast session, see Section 3) the Kalman filter can indeed be used in our context.

In Section 5 we restrict ourselves to the class of *linear filters* with the hope of relaxing some of the assumptions made in Section 4 for Kalman filtering theory to apply. The best filter is then a Wiener filter. We show that the Wiener filter can be computed for *any traffic regime* (as opposed to the Kalman filter in Section 4 that is derived in heavy-traffic regime) provided that on-times are exponentially distributed. Interestingly enough, both filters obtained in Sections 5 and 4 turn out to be identical. This observation thereby explains the good performance of the Kalman filter that we have observed under moderate and light traffic regimes (see Section 8).

In Section 6 we determine the optimal *first-order linear filter* for an *arbitrary* on-time distribution. We illustrate the approach in the case where the on-time distribution is hyperexponential.

The rest of the paper is organized as follows: motivation for this work is given in Section 2 and the multicast group model is introduced in Section 3. Estimators are obtained in Sections 4-6 for fixed parameters p and S; in Section 7 we give guidelines on how to choose these parameters so as to limit the number of ACKs travelling to the source, while in the meantime achieving a good quality of our estimators. The robustness of the estimators is addressed in Section 8. Extensions of our work are discussed in Section 9 and concluding remarks follow in Section 10.

2 Motivation

In order to best track the time-evolution of the multicast membership, we aim at developing an *unbiased* moving average estimator that would take advantage of previous estimates in an *optimal* way. We propose a mechanism in which the receivers probabilistically send "heartbeats" to the sender (hereafter called the source) in a periodic way: every S second each participant sends an ACK to the source with the probability p. Hence, the feedback implosion problem is addressed via a convenient choice of the reply (or ACK) probability p and of the "ACK time-interval" S. Note that S should be larger than the largest round-trip time between a receiver and the source. Times $t = nS$, for $n = 1, 2, \ldots$, will denote the end of each polling round, and Y_n will denote the total number of ACKs received at the nth observation step, i.e. in the interval of time $](n-1)S, nS]$. We denote by N_n the size of the multicast population at time nS and by $\hat{N}_n$ an estimator for N_n.

A naive approach to the estimation problem would consist in estimating N_n by the ratio Y_n/p, namely, by letting $\hat{N}_n = Y_n/p$. It has been shown in [2, Ch. 2] that this estimator behaves very poorly. This is partly due to the fact that it ignores the "history" of the membership process,

A less naive approach to filter out the noisy observations consists of using an exponential weighted moving average (EWMA) like the one used in [19]. A natural choice is

$$\hat{N}_n = \alpha \hat{N}_{n-1} + (1-\alpha) Y_n/p \tag{1}$$

which yields an (asymptotically) unbiased estimator, since $\mathbf{E}[\hat{N}_n] = \mathbf{E}[Y_n]/p = \mathbf{E}[N_n]$ in steady-state.

The difficulty in using the EWMA approach lies in the choice of the parameter α, as the performance of the estimator will in general be highly sensitive to this choice. This sensitivity is illustrated in Fig. 1, where the estimator has been computed on an audio trace for three different (but fairly close) values of α, namely, 0.95, 0.99 and 0.999. We

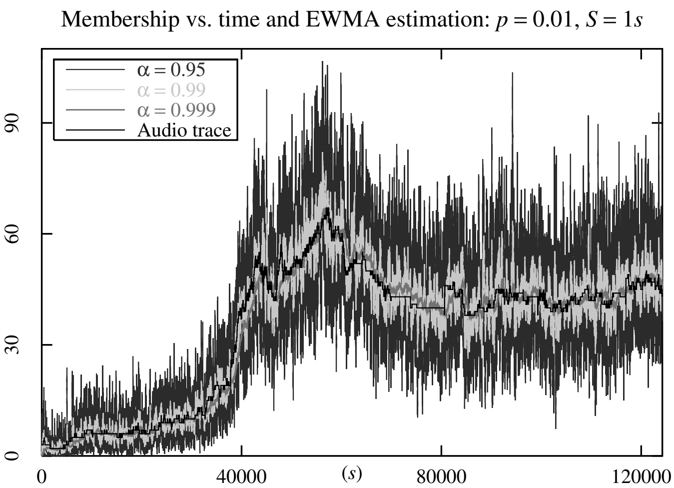

Figure 1: Membership evolution of a short audio session and EWMA estimation

can observe that the estimators computed for $\alpha = 0.95$ and $\alpha = 0.99$ are much more noisy than the estimator obtained for $\alpha = 0.999$, which appears to be very good. We are therefore left with the problem of selecting a "good" value for α, not an easy task since this value will typically be session dependent. Besides, there is no guarantee that an estimator based on the EWMA algorithm will be optimal in some sense (e.g. will minimize the mean square estimation error).

For these reasons, we will use another approach in the following and will rely on adaptive filter theory to construct optimal (to be made more precise) estimators.

Throughout the paper p and S are held fixed. In Section 7 we will give guidelines on how to select these parameters.

3 The multicast group model

In this section, we present the model for the multicast group. We consider a multicast group where participants join and leave at random times. Let T_i and $T_i + D_i$ be the join time and the leave time, respectively, of the ith participant. In the following, $D_i > 0$ is called the on-time of the ith participant and $\{D_i, i = 1, 2, \ldots\}$ is referred to as the on-time sequence.

Let $\tilde{N}(t)$ be the number of participants at time $t \geq 0$ or, equivalently, the size of the multicast audience at time t. We have

$$\tilde{N}(t) = \sum_{i=0}^{\tilde{N}(0)} D_i^{(r)} + \sum_{i=1}^{\infty} \mathbf{1}\{T_i \leq t < T_i + D_i\} \qquad (2)$$

where $\{D_i^{(r)}, i = 1, 2, \ldots, \tilde{N}(0)\}$ are the remaining on-times at $t = 0$ of participants, if any, which have joined the session before $t = 0$ and who are still connected at time $t = 0$ (with $D_0^{(r)} = 0$ by convention) and $\mathbf{1}\{E\}$ is the indicator function of any event E (i.e. $\mathbf{1}\{E\} = 1$ if the event E occurs and $\mathbf{1}\{E\} = 0$ otherwise).

Primarily for mathematical tractability we shall assume from now on that the join (arrival) process is Poisson with intensity $\lambda := 1/\mathbf{E}[T_{i+1} - T_i] > 0$ and that on-times form a renewal sequence of random variables (RVs) with common probability distribution $\Psi(x) = P(D_i < x)$ such that $0 < \mathbf{E}[D_i] < \infty$, further independent of the arrival times. In the following D will denote a generic RV with probability distribution $\Psi(x)$.

In the queueing terminology the process $\{\tilde{N}(t), t \geq 0\}$ is the occupation process (number of busy servers) in a $M/G/\infty$ queueing system with arrival rate λ and service times $\{D_i, i = 1, 2, \ldots\}$ [16].

For later use, we briefly review some results on the $M/G/\infty$ queue. In *steady-state*, the number N of busy servers is a Poisson RV with parameter $\rho := \lambda\, E[D]$, namely, $P[N = j] = \rho^j \exp(-\rho)/j!$. In particular, both the mean and the variance of the number of busy servers are equal to ρ. The autocovariance function of the *stationary version* of the process $\{\tilde{N}(t), t \geq 0\}$, denoted by $\{N(t), t \geq 0\}$, is given by [7, Equation (5.39)]

$$\mathrm{Cov}(N(t), N(t+h)) = \lambda \int_{|h|}^{\infty} P(D > u)\, \mathrm{d}u. \qquad (3)$$

From now on, we will only work with the stationary process $\{N(t), t \geq 0\}$, still for the sake of mathematical tractability. This is equivalent to assuming that when the tracking begins, the system has been operating sufficiently long with respect to session time durations (for instance, we can see on Fig. 1 that steady-state is reached after approximately 40,000 sec.). We have observed in our experiments (see [2, Ch. 2]) that the estimators we will develop in the forthcoming sections behave well even when the multicast population is not in steady-state at the beginning of the tracking (see Fig. 2 in Section 8) or when the steady-state assumption is violated during the entire estimation process (see Fig. 3 in Section 9).

We denote by $\{N_n, n = 0, 1, \ldots\}$ the process $\{N(t), t \geq 0\}$ sampled at times $t = 0, S, 2S, \ldots$, namely $N_n := N(nS)$.

Let $\mathrm{Cov}_X(\cdot)$ denote the autocovariance function of any second-order discrete-time stationary process $\{X_n, n = 0, 1, \ldots\}$. In the case where the on-times $\{D_i, i = 1, 2, \ldots\}$ are *exponentially* distributed with mean $E[D] = 1/\mu$, we have

$$\mathrm{Cov}_N(k) = \rho\, \gamma^{|k|}, \quad k = 0, \pm 1, \ldots \qquad (4)$$

with $\gamma := \exp(-\mu S)$.

Throughout, we will assume that

$$\sum_{k \geq 0} \mathrm{Cov}_N(k) < \infty \qquad (5)$$

thereby ruling out the situation where the on-times are heavy-tailed (e.g. Pareto distribution with shape parameter smaller than 2).

In the next three sections we derive three Mean-Square Error (MSE) *optimal* estimators for the size of the multicast audience at times nS ($n = 0, 1, \ldots$) under different sets of assumptions (exponential on-time distribution and heavy traffic regime in Section 4 by using a Kalman filter, exponential on-time distribution in Section 5 by using a Wiener filter and general on-time distribution in Section 6). In each case the optimality is defined with respect to a different class of filters (class of all measurable filters in Section 4, class of all linear filters in Section 5 and class of all first-order linear filters in Section 6).

A word on the notation used in this paper: $N(m, v)$ will denote a normal distribution with mean m and variance v and $X \sim N(m, v)$ will denote a RV with distribution $N(m, v)$; $\{a_n\}_n$ will stand for $\{a_n, n = 0, 1, \ldots\}$.

4 Optimal estimation using a Kalman filter

In this section, which reviews previous work published in [4], we derive an estimator of the size of the multicast audience at time nS by using Kalman filtering theory. This estimator will be obtained in heavy-traffic.

The heavy-traffic regime is obtained by "speeding up" the arrivals by a factor T or, equivalently, by assuming that the arrival intensity is now λT. We denote by $\{N_T(t), t \geq 0\}$ the occupation process in this new M/G/∞ queue with arrival rate λT. We will assume that the process $\{N_T(t), t \geq 0\}$ is stationary for all $T > 0$. Hence, $N_T(t)$ is a Poisson RV with parameter ρT for all $T > 0$, with $\rho := \lambda/\mu$ (see Section 3).

Let us introduce the normalized process $\{Z_T(t), t \geq 0\}$ defined by

$$Z_T(t) = \frac{N_T(t) - \rho T}{\sqrt{T}}, \quad t \geq 0. \qquad (6)$$

The process $\{Z_T(t), t \geq 0\}$ describes the fluctuations of $\{N_T(t), t \geq 0\}$ around its limiting trajectory ρT as $T \to \infty$. A nice feature of the process $\{Z_T(t), t \geq 0\}$ is that it converges to a diffusion process as $T \to \infty$ when the on-times are *exponentially distributed* RVs. More precisely, as $T \to \infty$ the (stationary) process $\{Z_T(t), t \geq 0\}$ converges in distribution to the Ornstein-Ühlenbeck process $\{X(t), t \geq 0\}$

given by [20, Theorem 6.14, page 155]

$$X(t) = e^{-\mu t} X(0) + \sqrt{2\lambda} \int_0^t e^{-\mu(t-u)}\, \mathrm{d}B(u), \qquad (7)$$

with $X(0) \sim N(0,\rho)$, where $\{B(t),\, t \geq 0\}$ is the standard Brownian motion. The Ornstein-Ühlenbeck process defined in (7) is a stationary ergodic Markov process, and its invariant distribution is a normal distribution with mean zero and variance ρ [15, page 358].

In the remainder of this section we will assume that the on-times $\{D_i, i = 1, 2, \ldots\}$ are exponentially distributed RVs.

We now show that the estimation problem can be reduced to a discrete filtering problem, to which discrete Kalman filtering theory applies. We first show that the process $\{X(t), t \geq 0\}$, sampled at discrete times $t = nS$, is governed by a linear stochastic difference equation; then, we show that the measurement equation at time nS is linear in the system state $X(nS)$.

4.1 System dynamics

From (7), we obtain, for $0 \leq s \leq t$, $X(t) = e^{-\mu(t-s)} X(s) + \sqrt{2\lambda} \int_s^t e^{-\mu(t-u)}\, \mathrm{d}B(u)$, from which it follows that

$$\xi_{n+1} = \gamma\, \xi_n + w_n, \quad n = 0, 1, \ldots \qquad (8)$$

where $\xi_n := X(nS)$, $\gamma := e^{-\mu S}$ and

$$w_n := \sqrt{2\lambda} \int_{nS}^{(n+1)S} e^{-\mu((n+1)S-u)}\, \mathrm{d}B(u).$$

The RVs $\{w_n\}_n$ are i.i.d. with

$$w_n \sim N(0, Q), \quad n = 0, 1, \ldots \qquad (9)$$

(see e.g. [6, page 17]) where Q is given by

$$\begin{aligned} Q &= 2\lambda\, \mathbf{E}\left[\int_{nS}^{(n+1)S} e^{-\mu((n+1)S-u)}\, \mathrm{d}B(u)\right]^2 \\ &= 2\lambda \int_{nS}^{(n+1)S} e^{-2\mu((n+1)S-u)}\, \mathrm{d}u = \rho\,(1-\gamma^2). \end{aligned}$$

Equation (8) establishes a linear stochastic difference equation relating the state of the limiting process $\{X(t),\, t \geq 0\}$ at consecutive polling instants nS and $(n+1)S$.

4.2 Measurement equation

Let ζ_n^i be the indicator function that receiver $i = 1, 2, \ldots, N_T(nS)$ has sent an ACK in the nth polling round, with $\zeta_n^i = 1$ if an ACK was sent by receiver i and $\zeta_n^i = 0$ otherwise. From the definition of the model it is seen that, conditioned on $N_T(nS)$, $\zeta_n^1, \ldots, \zeta_n^{N_T(nS)}$ are i.i.d. Bernoulli RVs with $\mathbf{E}[\zeta_n^i] = p$. The conditional expectation and variance of the number of ACKs $Y_n = \sum_{i=1}^{N_T(nS)} \zeta_n^i$ received by the source at time nS are then given by $N_T(nS)\, p$ and $N_T(nS)\, p(1-p)$, respectively. We define our normalized measurement equation as

$$M_T(nS) = \frac{Y_n - p\rho T}{\sqrt{T}}, \quad n = 0, 1, \ldots. \qquad (10)$$

which, with the help of (6), can be rewritten as

$$M_T(nS) = p\, Z_T(nS) + V_T(nS), \qquad (11)$$

where

$$V_T(nS) := \frac{Y_n - N_T(nS)p}{\sqrt{T}}. \qquad (12)$$

The next step is to let $T \to \infty$ in (11). The following proposition is proved in [4].

Proposition 4.1 *There exist i.i.d. RVs $\{v_n,\, n = 0, 1, \ldots\}$ with*

$$v_n \sim N(0, R), \quad n = 0, 1, \ldots \qquad (13)$$

where $R := \rho\, p\,(1-p)$, independent of $\{w_n\}_n$, such that $\{v_k, k = n, n+1, \ldots\}$ is independent of $\{\xi_k, k = 0, 1, \ldots, n\}$ for $n = 0, 1, \ldots$, and such that $(Z_T(nS), V_T(nS))$ converges weakly to (ξ_n, v_n) as $T \to \infty$. ♦

We deduce from Proposition 4.1 that $M_T(nS)$ defined in (10) converges weakly as $T \to \infty$ to a RV m_n such that

$$m_n = p\xi_n + v_n, \quad n = 0, 1, \ldots. \qquad (14)$$

4.3 Deriving the filter parameters

Equations (8) and (14) represent the equations of a discrete time linear filter, for which we can compute the optimal estimator. Throughout we shall assume that the Gaussian initial condition ξ_0, the signal noise sequence $\{w_n\}_n$ and the observation noise sequence $\{v_n\}_n$ are all mutually independent.

Let $\hat{\xi}_n$ be an estimator of ξ_n, and denote by $\epsilon_n = \xi_n - \hat{\xi}_n$ the estimation error. The estimator that minimizes the mean square of the estimation error is given by the following Kalman filter (see e.g. [23, page 347]), which, in its stationary version, has the following simple recursive structure:

$$\begin{aligned} P &= \Big(\big(\gamma^2 P + Q\big)^{-1} + p^2/R\Big)^{-1} & (15) \\ K &= Pp/R & (16) \\ \hat{\xi}_n &= \gamma\hat{\xi}_{n-1} + K\Big(m_n - p\Big(\gamma\hat{\xi}_{n-1}\Big)\Big) & (17) \end{aligned}$$

for $n = 1, 2, \ldots$, with $\hat{\xi}_0 = \mathbf{E}[\xi_0] = 0$ and where constants γ, R and Q have been defined earlier in the section.

The Ricatti equation (15) has a unique positive solution P given by

$$\begin{aligned} P &= -\frac{Qp^2 + R\left(1-\gamma^2\right)}{2p^2\gamma^2} \\ &\quad + \frac{\sqrt{(Qp^2 + R\,(1-\gamma^2))^2 + 4p^2\gamma^2 RQ}}{2p^2\gamma^2}. \qquad (18) \end{aligned}$$

P gives the (stationary) variance of the estimation error. From (18) and (16) we find that the gain K is given by

$$K = \frac{-(1-\gamma^2) + \sqrt{(1-\gamma^2)(1-\gamma^2(1-2p)^2)}}{2\gamma^2 p(1-p)}. \quad (19)$$

Recall that $\epsilon_n \sim N(0, P)$ for every n and that ϵ_n is independent of the observation m_n [24, page 240].

4.4 Membership size estimation

We now return to our original estimation problem, namely, the derivation of an estimator $(\hat{N}_n)$ for the size of the multicast group at time nS (i.e. $N_T(nS)$). Recall that the process $\{N_T(t),\, t \geq 0\}$ describes the number of busy servers in a stationary $M/M/\infty$ queue with arrival rate λT and service rate μ. Motivated by (6), we define $\hat{N}_n$ as follows:

$$\hat{N}_n = \hat{\xi}_n \sqrt{T} + \rho T \quad (20)$$

with $\hat{\xi}_n$ given in (17). Combining (17), (10) and (20), we find the following first-order linear equation

$$\hat{N}_n = \gamma(1-Kp)\hat{N}_{n-1} + K\, Y_n + \rho T(1-\gamma)(1-Kp). \quad (21)$$

Starting with $\mathbf{E}[\hat{\xi}_0] = 0$ it is seen from (17) and (14) that $\mathbf{E}[\hat{\xi}_n] = 0$ which in turn implies from (20) that $\mathbf{E}[\hat{N}_n] = \rho T = \mathbf{E}[N_T(nS)]$. This shows that $\hat{N}_n$ is an unbiased estimator. On the other hand, $\mathrm{Var}((N_n - \hat{N}_n)\sqrt{T}) = \mathrm{Var}(Z_T(nS) - \hat{\xi}_n)$ from (6) and (20); we conjecture that, as $T \to \infty$, the latter quantity converges to P, the variance of the estimation error ϵ_n in heavy-traffic.

The estimation algorithm is summarized below (ρT, μ and S are assumed to be known):

`Initialization step:`

$\hat{N}_0 = \rho T$ (i.e. $\hat{\xi}_0 = 0$), $\gamma = \exp(-\mu S)$ and set gain K as given in (19).

n`th observation step:`

Y_n = number of ACKs received in interval of time $](n-1)S, nS]$ and compute $\hat{N}_n$ as in (21).

Guidelines for choosing parameters p and S are given in Section 7; a procedure for estimating parameters ρT (expected number of participants) and $1/\mu$ (expected on-time) is discussed in Section 9.

Remark 4.1 *The autoregressive equation in (21) does not exhibit the same form as the one in (1) as it further has a constant term $\rho T(1-\gamma)(1-Kp)$. In other words, if we had computed the optimal α in (1) under the assumptions considered in Section 3, we would not have obtained the optimal estimator.*

5 Optimal estimation using a Wiener filter

In the previous section we have derived a filter that is MSE-optimal among all measurable filters, provided that the system evolves in heavy-traffic (i.e. very large multicast audience) and that on-times are exponentially distributed.

In this section we will derive a (Wiener) filter that is MSE-optimal among all linear filters, under the only assumption that on-times are exponentially distributed.

The first step is to replace processes $\{N_n\}_n$, $\{\hat{N}_n\}_n$ and $\{Y_n\}_n$ by their centered (zero mean) versions $\{\nu_n\}_n$, $\{\hat{\nu}_n\}_n$ and $\{y_n\}_n$, respectively. We already know that $\mathbf{E}[N_n] = \rho$ (see Section 3). On the other hand,

$$\mathbf{E}[Y_n] = \mathbf{E}[\mathbf{E}[Y_n \mid N_n]] = \mathbf{E}[p\, N_n] = p\rho. \quad (22)$$

Taking $\nu_n := N_n - \rho$, $\hat{\nu}_n := \hat{N}_n - \rho$ and $y_n := Y_n - p\rho$ will therefore ensure that $\mathbf{E}[\nu_n] = \mathbf{E}[\hat{\nu}_n] = \mathbf{E}[y_n] = 0$.

Wiener filtering theory identifies the MSE-optimal *linear* filter, from which we get the following MSE-optimal estimator [13]

$$\nu_n = \sum_{k=0}^{\infty} h_{o,k} y_{n-k}$$

where the so-called optimal impulse response $\{h_{o,n}\}_n$ satisfies the Wiener-Hopf equation

$$\sum_{m=0}^{\infty} h_{o,m} \mathrm{Cov}_y(k-m) = \mathrm{Cov}_{\nu y}(k), \quad k = 0, 1, \ldots. \quad (23)$$

In (23) $\mathrm{Cov}_y(k)$ denotes the autocorrelation of the filter input (the measurements) $\{y_n\}_n$ and $\mathrm{Cov}_{\nu y}(k) = \mathbf{E}[\nu_{n-k}\, y_n]$ denotes the cross-correlation function of processes $\{\nu_n\}_n$ and $\{y_n\}_n$.

Therefore, all what we have to do is to compute $\mathrm{Cov}_y(k)$ and $\mathrm{Cov}_{\nu y}(k)$ and then to solve (23).

We can express $\mathrm{Cov}_y(k)$ and $\mathrm{Cov}_{\nu y}(k)$ in terms of $\mathrm{Cov}_\nu(k)$ as follows:

$$\mathrm{Cov}_y(k) = p^2 \mathrm{Cov}_\nu(k) + \mathbf{1}\{k=0\}\rho p(1-p) \quad (24)$$
$$\mathrm{Cov}_{\nu y}(k) = p\, \mathrm{Cov}_\nu(k) \quad (25)$$

where we have used the identity $\mathrm{Cov}_\nu(k) = \mathrm{Cov}_N(k)$.

One way of solving the Wiener-Hopf equation (23) is instantiated in the *prewhitening approach* [13, page 81] whose steps are given below: for $|z| = 1$

- The power spectrum of the input signal $\{y_n\}_n$, $S_y(z) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}_y(k) z^{-k}$, is factorized as

  $$S_y(z) = \sigma^2 G(z) G(z^{-1}), \quad (26)$$

  where σ^2 is a constant and $G(z)$ is the part of $S_y(z)$ having all its zeros and poles *inside* the unit circle (therefore $G(z^{-1})$ is the part of $S_y(z)$ having all its zeros and poles *outside* the unit circle).

- The cross-power spectrum between $\{\nu_n\}_n$ and $\{y_n\}_n$, $S_{\nu y}(z) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}_{\nu y}(k) z^{-k}$, is then divided by $G(z^{-1})$. Expanding this ratio into fractions, then taking the fractions with zeros and poles inside the unit circle and dividing the resulting fractions by σ^2 gives $H'_o(z) = (1/\sigma^2)\left[S_{\nu y}(z)/G(z^{-1})\right]_+$.

- The transfer function of the Wiener Filter, $H_o(z)$, is formed by multiplying $H'_o(z)$ by $1/G(z)$.

- Inverting the transfer function of the optimal filter, $H_o(z) = H'_o(z)/G(z) = \sum_{k=0}^{\infty} h_{o,k} z^{-k}$, back into the time domain yields the desired recurrence between $\hat{\nu}_n$ and y_n and, subsequently, between the non-centered processes $\hat{N}_n$ and Y_n.

The success of the prewhitening approach rests on the ability to factorize the power spectrum of the original input signal $\{y_n\}_n$ as in (26). Unfortunately, we were able to perform this canonical factorization only when the underlying model is the $M/M/\infty$ queue (i.e. "exponential" on-times), which is illustrated in Section 5.1.

5.1 Application to the $M/M/\infty$ model

To compute the transfer function of the filter, we need to find expressions for $S_y(z)$ and $S_{\nu y}(z)$. Let us first determine $S_y(z)$. By using (24) and (4) together with the property $\mathrm{Cov}_N(k) = \mathrm{Cov}_\nu(k)$, we find

$$\mathrm{Cov}_y(k) = \begin{cases} p^2 \rho \gamma^{|k|}, & \text{for } k \neq 0 \\ p\rho, & \text{for } k = 0. \end{cases}$$

Since $\gamma = \exp(-\mu S) < 1$ and $|z| = 1$, the z-transform of $\mathrm{Cov}_y(k)$ is

$$S_y(z) = p\rho \frac{\gamma(p-1)z^2 + (1+\gamma^2(1-2p))z + \gamma(p-1)}{z(1-\gamma z)(1-\gamma z^{-1})}.$$

The second-order polynomial in the variable z in the numerator has two positive real roots given by $r < 1$ and $1/r > 1$, with

$$r = \frac{1+\gamma^2(1-2p) - \sqrt{(1-\gamma^2)[1-\gamma^2(1-2p)^2]}}{2\gamma(1-p)}.$$

Hence $S_y(z) = \sigma^2\, G(z)\, G(z^{-1})$ with $\sigma^2 := \gamma\rho p(1-p)/r$, and $G(z) := (1 - rz^{-1})/(1-\gamma z^{-1})$. We now compute $S_{\nu y}(z)$. From (25) and (4) we find $\mathrm{Cov}_{\nu y}(k) = p\rho\gamma^{|k|}$ so that

$$S_{\nu y}(z) = \frac{p\rho(1-\gamma^2)}{(1-\gamma z)(1-\gamma z^{-1})}.$$

The transfer function $H'_o(z)$ is given by

$$H'_o(z) = \frac{1}{\sigma^2}\left[\frac{S_{\nu y}(z)}{G(z^{-1})}\right]_+ = \frac{r(1-\gamma^2)}{\gamma(1-p)(1-\gamma r)(1-\gamma z^{-1})}$$

and the transfer function $H_o(z)$ of the optimal filter takes here the simple form

$$H_o(z) = \frac{r(1-\gamma^2)}{\gamma(1-p)(1-\gamma r)(1-rz^{-1})} = \frac{B}{1-Az^{-1}}$$

where $A = r$ and

$$\begin{aligned} B &= \frac{r(1-\gamma^2)}{\gamma(1-p)(1-\gamma r)} \\ &= \frac{-(1-\gamma^2) + \sqrt{(1-\gamma^2)(1-\gamma^2(1-2p)^2)}}{2\gamma^2 p(1-p)}. \end{aligned} \quad (27)$$

The impulse response of this linear filter is given by the *first-order* recurrence relation [13] $\hat{\nu}_n = A\hat{\nu}_{n-1} + By_n$, with $\hat{\nu}_n$ the estimator of ν_n. We now return to the original processes $\{\hat{N}_n\}_n$ and $\{Y_n\}_n$, to finally obtain the optimal linear filter

$$\hat{N}_n = A\hat{N}_{n-1} + BY_n + \rho(1-A-pB). \quad (28)$$

It is interesting to compare this filter with the Kalman filter derived in Section 4 (see (21), in which the filter gain K is given in (19)). Looking at (27) and (19), we can see that they are exactly the same. Developing the coefficient of $\hat{N}_{n-1}$ in (21), we obtain $\gamma(1-Kp) = A$. It remains to compare the constant terms in (21) and (28). Recall that ρT in Section 4 denotes the actual average number of receivers which is simply denoted by ρ in the present section. Developing the constant terms in both linear filters we find $(1-\gamma)(1-Kp) = 1 - A - pB$. We have therefore shown that the filters returned by both the Kalman theory and the Wiener theory are identical.

This result is not so surprising, since both the Kalman filter and the Wiener filter are MSE- optimal among the class of linear filters. The key point is that the Kalman filter used in Section 4 was derived under a heavy traffic assumption, while the Wiener filter computed in the present section holds for any value of the model parameters λ and μ. On the other hand, the Wiener filter is only optimal among all *linear* filters whereas the Kalman filter in Section 4 is optimal among all measurable filters.

We conclude this section by computing the mean square error $\epsilon_{min} := \mathbf{E}[(N_n - \hat{N}_n)^2]$ of our estimator. It is known that [13] $\epsilon_{min} = \sum_{k=1}^{M} \mathrm{Res}[F(z), z_k]$ with $F(z) := 1/z(S_\nu(z) - H_o(z)S_{\nu y}(z^{-1}))$ where $z_1, \ldots, z_M$ are the poles (if any) of the function $F(z)$ inside the unit circle. The notation $\mathrm{Res}[F(z), z_k]$ stands for the residue of $F(z)$ at point $z = z_k$. Specializing $F(z)$ to the values of $S_\nu(z)$, $S_{\nu y}(z)$, $H_o(z)$ found earlier, yields $F(z) = \dfrac{\rho(1-\gamma^2)((1-Bp)z - A)}{(1-\gamma z)(z-\gamma)(z-A)}$. This function has two poles inside the unit circle which are located at $z = A$ and $z = \gamma$; the residues of $F(z)$ at these poles are given by $-\rho pAB(1-\gamma^2)/[(1-\gamma A)(A-\gamma)]$ and $\rho[1 + pB\gamma/(A-\gamma)]$, respectively. Summing up these residues gives $\epsilon_{min} = \rho\left(1 - \frac{Bp}{1-\gamma A}\right)$. By using the expressions of A and B, we finally obtain

$$\epsilon_{min} = \rho \frac{-(1-\gamma^2) + \sqrt{(1-\gamma^2)(1-\gamma^2(1-2p)^2)}}{2\gamma^2 p}. \quad (29)$$

This expression for ϵ_{min} can be used to tune the parameters p and γ or equivalently S (see Section 7).

6 The optimal first-order linear filter

The theory reported in Section 5 applies to any on-time distribution $\Psi(x)$ such that (5) holds. However, it is not easy to identify the function $G(z)$ that appears in the canonical factorization of the spectrum $S_y(z)$ (see (26)) and thereby the optimal filter, except when the on-times are exponentially distributed RVs. As already pointed out, we would like to develop an estimator under the only assumptions introduced in Section 3 (namely Poisson join times and generally distributed on-times such that (5) holds).

In this section, we will use a least square estimation method to determine the first-order linear filter that minimizes the mean square error. Observe that, unlike the Wiener filter, the proposed approach will not return the optimal filter among all linear filters but simply the optimal linear filter among all first-order linear filters. We will illustrate this approach at the end of this section in the case where $\Psi(x)$ is a hyperexponential distribution. Recall the definition of the centered stationary processes $\{\nu_n\}_n$, $\{\hat{\nu}_n\}_n$ and $\{y_n\}_n$ introduced in Section 5.

The methodology is simple: we want to find constants $A \in (0,1)$ and B such that $\epsilon := \mathbf{E}[(\nu_n - \hat{\nu}_n)^2]$ is minimized when the process $\{\hat{\nu}_n\}_n$ satisfies the following first-order recurrence relation

$$\hat{\nu}_n = A\hat{\nu}_{n-1} + By_n. \tag{30}$$

In steady-state we have

$$\hat{\nu}_n = B\sum_{k=0}^{\infty} A^k y_{n-k}. \tag{31}$$

The mean square error ϵ is equal to $\epsilon = \mathbf{E}[\nu_n^2] - 2\mathbf{E}[\nu_n\hat{\nu}_n] + \mathbf{E}[\hat{\nu}_n^2]$. Therefore, we need to compute three terms to evaluate ϵ. First, $\mathbf{E}[\nu_n^2] = \mathbf{E}[(N_n-\rho)^2] = \rho$. Second, using (31) and (25) yields $\mathbf{E}[\nu_n\hat{\nu}_n] = pB\sum_{k=0}^{\infty} A^k \mathrm{Cov}_\nu(k) = pBg(A)$ where

$$g(z) := \sum_{k=0}^{\infty} z^k \mathrm{Cov}_\nu(k). \tag{32}$$

Third, squaring both sides of (30) and then taking the expectation yields $\mathbf{E}[\hat{\nu}_n^2] = \left(\frac{B}{1-A^2}\right)(2A\mathbf{E}[\hat{\nu}_{n-1}y_n] + B\mathbf{E}[y_n^2])$. We know that $\mathbf{E}[y_n^2] = \mathrm{Cov}_y(0) = \rho p$ (see (24)) and from (31), (24) and $\mathrm{Cov}_\nu(0) = \rho$ we have $\mathbf{E}[\hat{\nu}_{n-1}y_n] = Bp^2\,(g(A) - \rho)/A$. We finally obtain $\mathbf{E}[\hat{\nu}_n^2] = \left(\frac{pB^2}{1-A^2}\right)(2pg(A) + \rho(1-2p))$. Having computed $\mathbf{E}[\nu_n^2]$, $\mathbf{E}[\nu_n\hat{\nu}_n]$ and $\mathbf{E}[\hat{\nu}_n^2]$, we can write the mean square error as follows

$$\epsilon = \rho - 2pBg(A) + \left(\frac{pB^2}{1-A^2}\right)(2pg(A) + \rho(1-2p)). \tag{33}$$

Observe that the power series $g(z)$ converges for $|z| < 1$ (since $k \to \mathrm{Cov}_\nu(k)$ is non-increasing) and is therefore differentiable for $|z| < 1$. We will denote by $g'(z)$ its derivative.

In order to minimize ϵ, $A \in (0,1)$ and B must be the solution of the following system of equations:

$$\begin{cases} \dfrac{\partial\epsilon}{\partial A} = \dfrac{2pB}{1-A^2}\Bigg(AB\left[\dfrac{2pg(A)+\rho(1-2p)}{1-A^2}\right] \\ \qquad\qquad +g'(A)(pB-(1-A^2))\Bigg) = 0 \\ \dfrac{\partial\epsilon}{\partial B} = 2p\left(B\left[\dfrac{2p(g(A)-\rho)+\rho}{1-A^2}\right] - g(A)\right) = 0. \end{cases}$$

The second equation gives

$$B = \frac{g(A)(1-A^2)}{2p(g(A)-\rho)+\rho}. \tag{34}$$

Substituting this value of B into the first equation shows that A must satisfy

$$\begin{aligned} &Ag(A)(2p(g(A)-\rho)+\rho) \\ &\quad -g'(A)(1-A^2)(p(g(A)-\rho)+\rho(1-p)) = 0 \end{aligned}$$

If this equation has a unique solution $A \in (0,1)$, then substituting this value of A into (34) will give the optimal pair (A,B). Proposition 6.1 shows that this is indeed the case (see [3] for a proof).

Proposition 6.1 *Define $f(x) := (2p(g(x)-\rho)+\rho)xg(x) - (p(g(x)-\rho)+\rho(1-p))(1-x^2)g'(x)$, where $g(x)$ is given in (32). If $g'(x) > 0$ for $x \in [0,1)$, then $f(x)$ has a unique zero in $[0,1)$.* ♦

The reader can check that the filter defined in (30) with the optimal pair (A,B) is the same as the Wiener filter found in Section 5.1 when the on-times are exponentially distributed.

6.1 Application to the $M/H_L/\infty$ model

We now illustrate the approach developed in this section by considering the situation where on-times follow a hyperexponential distribution. More precisely, we assume that

$$\Psi(x) = 1 - \sum_{l=1}^{L} p_l e^{-\mu_l x} \tag{35}$$

with $0 < p_l < 1$, $l = 1,2,\ldots,L$, and $\sum_{l=1}^{L} p_l = 1$. In this setting, the underlying queueing model can be seen as L independent $M/M/\infty$ queues in parallel. The arrival rate to queue l is $p_l\lambda$ and the service rate is μ_l. Define $\gamma_l := \exp(-\mu_l S)$, $\rho_l := p_l\lambda/\mu_l$ so that $\rho = \sum_{l=1}^{L}\rho_l$. The autocovariance function of the process $\{\nu_n, n = 0,1,\ldots\}$ is equal to $\mathrm{Cov}_\nu(k) = \sum_{l=1}^{L}\rho_l\gamma_l^{|k|}$ so that $g(A) = \displaystyle\sum_{l=1}^{L}\frac{\rho_l}{1-A\gamma_l}$.

Numerical example [1]

$L = 2$, $p = 0.0106$ and $S = 2.5s$. Also

$$\begin{array}{lllllllll} 1/\mu_1 & = & 3897s, & \rho_1 & = & 19.5, & \gamma_1 & = & 0.999359 \\ 1/\mu_2 & = & 480061s, & \rho_2 & = & 75.1, & \gamma_2 & = & 0.999995 \\ 1/\mu & = & 18316s, & \rho & = & 94.7. & & & \end{array}$$

The optimal first-order filter is

$$\hat{N}_n = 0.99879456\,\hat{N}_{n-1} + 0.10720289\,Y_n + 0.006540864.$$

For comparison, the Wiener filter found in Section 5.1 (for exponential on-times) is

$$\hat{N}_n = 0.99828589\,\hat{N}_{n-1} + 0.14885344\,Y_n + 0.012900081.$$

7 Guidelines on choosing p and S

A "good" pair (p, S) should (i) limit the feedback implosion while at the same time (ii) achieve a good quality of the estimator. Of course (i) and (ii) are antinomic and therefore a trade-off must be found. This trade-off will be formalized as follows: we want to select a pair (p, S) so that the mean number of ACKs generated every S seconds (see (22)) and the relative error of the variance of the estimator (denoted as η) are bounded from above by given constants, namely

$$\begin{cases} \mathbf{E}[Y_n] = p\rho \le \alpha \\ \eta = \dfrac{\mathrm{Var}(N_n) - \mathrm{Var}(\hat{N}_n)}{\mathrm{Var}(N_n)} \le \beta. \end{cases} \tag{36}$$

When $\hat{N}_n$ is optimal among all linear filters, then $\mathrm{Var}(N_n) - \mathrm{Var}(\hat{N}_n) = \mathbf{E}[(N_n - \hat{N}_n)^2]$ and η becomes the "normalized mean square error" [14, page 202]. Optimality was shown for the $M/M/\infty$ queue, therefore $\eta = \epsilon_{min}/\rho$ with ϵ_{min} given in (29).

For given constants α and β, it is easy to solve the constrained optimization problem defined in (36), provided that η is known. For the $M/M/\infty$ model, where ϵ_{min} is given in (29), we find that $p = \alpha/\rho$ and that S, or equivalently γ, is the unique positive solution of the equation $\epsilon_{min} = \rho\beta$. The problem now is to choose constants α and β so that conditions (i) and (ii) are satisfied. We have found in our experiments that α in the range $[0.5, 1]$ and $\beta \le 0.15$ give satisfactory results.

We conclude this section with general remarks on how to adapt the parameters p and S to important variations in the membership. The estimation schemes in Sections 4.3, 5.1 and 6.1 have been obtained under the assumption that parameters p and S are fixed. However, the filters therein constructed can still be used if p and/or S change over time, provided that these modifications do not prevent the system to be in steady-state most of the time. In that setting, a new filter will have to be recomputed after each modification. Such a modification can be carried out each time the number of ACKs received during a given period of time significantly deviates from the current expectation (i.e. $p\rho$).

8 Validation with real video traces

In this section we apply the estimators developed in Sections 5.1 and 6.1 to four traces of real video sessions. Two types of estimators will be used: the estimator – denoted as $\hat{N}_n^E$ – found in (28) when the population is modeled as an $M/M/\infty$ queue; the estimator – denoted as $\hat{N}_n^{H_2}$ – derived in Section 6.1 in the case where join times are Poisson and on-times have a 2-stage hyperexponential distribution ($M/H_2/\infty$ model).

The objective is twofold: we want to investigate the quality of both estimators when compared to real life conditions, and we want to identify the best one. We have collected four MBone traces – denoted $video_i, i = 1, \ldots, 4$ – between August 2001 and September 2001 using the *MListen* tool [1]. Each trace corresponds to a long-lived video session (see duration of each session in Table 1, where the superscript "d" stands for "days") and records the pair (T_i, D_i) for each participant in the session. We have run both algorithms (estimators) on each trace. For each trace, we have identified the parameters of the $M/M/\infty$ model (parameters λ and μ, or equivalently parameters ρ and μ) and of the $M/H_2/\infty$ model (parameters ρ, μ_1, μ_2, p_1 and $p_2 = 1 - p_1$). The values of these parameters are reported in columns 3–8 in Table 1. Parameters p and S have been chosen by following the guidelines presented in Section 7. Values of these parameters are listed in columns 9–10 in Table 1. The performance of estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ are reported in Tables 2 and 3.

Table 2 reports several order statistics (columns 3–7) and the sample mean of the relative error $\frac{|N_n - \hat{N}_n|}{N_n}$ (column 2), where $\hat{N}_n$ is either $\hat{N}_n^E$ or $\hat{N}_n^{H_2}$. All results are expressed in percentages. The first observation is that both estimators perform reasonably well. The sample mean of the relative error is always less than 6.82% and is as low as 3.79%; when averaging over all experiments, this sample mean is less than 4.5% for both $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ (see last two rows). The second observation is that no scheme is uniformly better than the other one over an entire session but their sample means are very close to each other (see column 2). For instance, $\hat{N}_n^E$ performs better than $\hat{N}_n^{H_2}$ regarding the 90th and the 95th percentiles whereas the result is reversed regarding the 25th percentile. It looks like the relative error on $\hat{N}_n^{H_2}$ is empirically more dispersed around its mean than is the relative error on $\hat{N}_n^E$, and has a longer tail.

Table 3 reports the sample mean and the sample variance of the error $N_n - \hat{N}_n$. In the 4th column, we list the theoretical variance. It is given by ϵ_{min} for $\hat{N}_n^E$ (see (29)) and by ϵ for $\hat{N}_n^{H_2}$ (see (33)). The expected average $\mathbf{E}[N_n - \hat{N}_n]$ is zero in both approaches. Both estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$

[1] The values of the parameters come from the trace called $video_1$ investigated in Section 8.

Table 1: Parameter identification

Trace	Session lifetime	ρ	$1/\mu$	$1/\mu_1$	$1/\mu_2$	p_1	p_2	p	S	α	β
$video_1$	$3^d\,13^h\,33^m\,20^s$	94.7	18316	3897	480061	0.97	0.03	0.011	2.5	1.0	0.15
$video_2$	$11^d\ \ 1^h\,46^m\ \ 8^s$	14.1	16476	1	226498	0.93	0.07	0.034	3.2	0.5	0.1
$video_3$	$50^d\,22^h\,13^m\,20^s$	8.1	66823	1	900854	0.93	0.07	0.062	20.0	0.5	0.1
$video_4$	$29^d\,16^h\,43^m\,13^s$	17.9	83390	1	473268	0.82	0.18	0.028	10.0	0.5	0.1

Table 2: Mean and percentiles of relative error $|N_n - \hat{N}_n|/N_n$

Trace	Estimator	Mean	25	50	75	90	95
$video_1$	$\hat{N}_n^E$	6.82	1.09	2.42	5.25	11.5	19.4
	$\hat{N}_n^{H_2}$	6.12	1.08	2.55	6.31	13.5	20.6
$video_2$	$\hat{N}_n^E$	4.19	1.41	3.08	5.43	8.66	11.9
	$\hat{N}_n^{H_2}$	4.12	0.98	2.14	4.41	8.78	12.6
$video_3$	$\hat{N}_n^E$	4.20	1.55	3.26	5.71	8.71	11.0
	$\hat{N}_n^{H_2}$	3.98	1.07	2.36	4.83	9.35	12.6
$video_4$	$\hat{N}_n^E$	3.79	1.23	2.57	4.51	7.50	11.0
	$\hat{N}_n^{H_2}$	4.06	1.02	2.21	4.39	8.98	14.7
All	$\hat{N}_n^E$	4.44	1.33	2.88	5.22	8.60	12.0
	$\hat{N}_n^{H_2}$	4.34	1.02	2.26	4.73	9.61	14.2

Table 3: Empirical mean and variance of the error $N_n - \hat{N}_n$

Trace	Estimator	Mean	Variance	ϵ_{min}, ϵ	η
$video_1$	$\hat{N}_n^E$	−0.112	12.664	13.942	0.147
	$\hat{N}_n^{H_2}$	−0.047	12.851	12.120	
$video_2$	$\hat{N}_n^E$	0.006	0.495	1.407	0.099
	$\hat{N}_n^{H_2}$	0.019	0.785	0.396	
$video_3$	$\hat{N}_n^E$	0.037	0.207	0.737	0.091
	$\hat{N}_n^{H_2}$	0.019	0.229	0.208	
$video_4$	$\hat{N}_n^E$	0.052	0.911	1.566	0.087
	$\hat{N}_n^{H_2}$	0.065	1.423	0.676	

have almost no bias (see column 2), and their empirical variances closely match the theoretical ones given by ϵ_{min} and ϵ, respectively. It is of interest to point out that for the 4 traces studied, ϵ, the theoretical mean square error provided by $\hat{N}_n^{H_2}$, is smaller than ϵ_{min}, the theoretical mean square error provided by $\hat{N}_n^E$ (however, this result is reversed if we consider the empirical mean square errors). Thus, $\hat{N}_n^{H_2}$ is more efficient (an estimator is said to be more efficient if it has a smaller variance) than $\hat{N}_n^E$ (again, $\hat{N}_n^E$ is empirically more efficient than $\hat{N}_n^{H_2}$). The last column provides the relative error on $\mathrm{Var}(\hat{N}_n^E)$, called η $(= \epsilon_{min}/\rho)$ in Section 7. Notice that $\eta < \beta$ (β is given in column 12 in Table 1).

Fig. 2 displays the variations of membership for session $video_1$ (which presents the highest variations in N_n) together with the estimates returned by $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$. Fig. 2(a) displays three curves: the collected video trace, the estimation returned by $\hat{N}_n^E$, labeled "Exponential", and the estimation returned by $\hat{N}_n^{H_2}$, labeled "Hyperexponential". It appears that $\hat{N}_n^E$ follows better N_n during periods of high variations whereas $\hat{N}_n^{H_2}$ is slightly closer to N_n during flat periods.

Both estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ have been derived under some specific and restrictive assumptions: Poisson join times for both of them, exponential (resp. 2-stage hyperexponential) on-times for the first (resp. second) one. It is interesting to know whether or not these assumptions were violated in each session $video_i$, $i = 1, \ldots, 4$. We have therefore carried out a statistical analysis of each trace in order to determine the nature of their join time process and of their on-time sequence. As shown in Table 4 and Fig. 2, parts (b) and (c), neither is the join time process Poisson nor are on-times exponentially distributed (or hyperexponentially distributed), for any of the traces. The inter-join times and the on-times appear to follow subexponential distributions (Lognormal and Weibull distributions), a situation quite different from the assumptions under which the estimators have been obtained. Despite these significant differences, the estimators behave well and therefore show a good robustness to assumption violations.

In summary, both estimators perform very well when applied to real traces and are robust to significant deviations from their (theoretical) domain of validity. Estimator $\hat{N}_n^{H_2}$ returns the best global performance for the relative error criterion, but does not track high fluctuations as well as $\hat{N}_n^E$. Overall, we have found that $\hat{N}_n^E$ is a good estimator, both in terms of its performance and its usability since it only requires the knowledge of two parameters: ρ and μ.

9 Estimating parameters ρ and μ

The main pending issue concerns the knowledge of parameters ρ and μ (or equivalently any two parameters among ρ, λ and μ, since $\rho = \lambda/\mu$ in steady-state). When these parameters are not known, the source should estimate them. Again, the source could estimate any two parameters among ρ, λ and μ and infer the third one.

One possible way of estimating λ is to let a newly arrived receiver send a "hello" message to the source with a certain (constant) probability q (q should be small enough to avoid overwhelming the source with hellos). The source would then use the arrival time t_m of the mth hello to estimate λ.

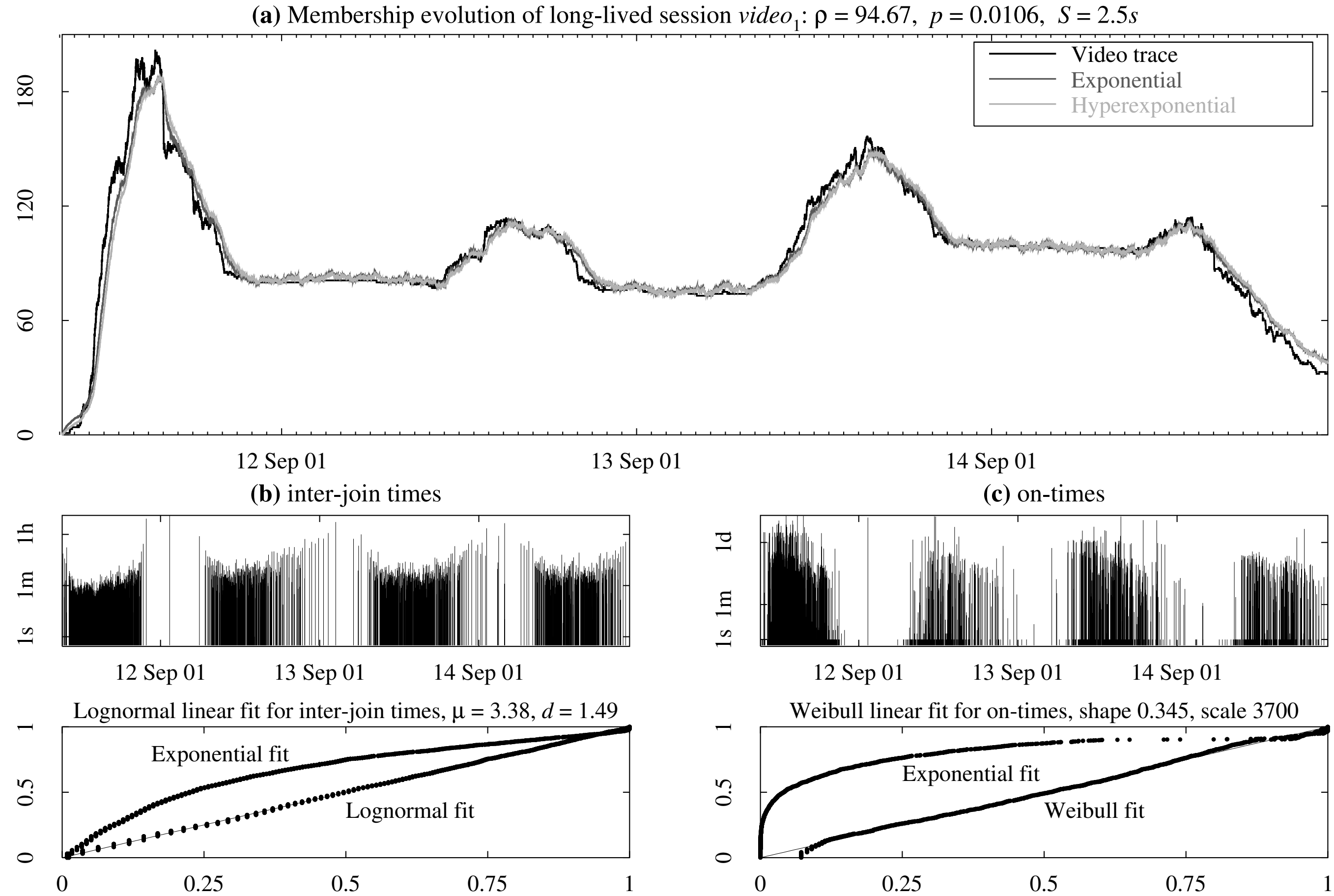

Figure 2: Membership estimation of session $video_1$ and corresponding probability plots

Table 4: Distributions that best fitted into the inter-arrivals and on-times sequences

Trace	Best fit for inter-arrivals sequence	Best fit for on-times sequence
$video_1$	Lognormal with $\mu = 3.38$, $d = 1.49$	Weibull with shape 0.35, scale 3700
$video_2$	Lognormal with $\mu = 5.20$, $d = 1.68$	Weibull with shape 0.26, scale 1400
$video_3$	Weibull with shape 0.65, scale 3500	Lognormal with $\mu = 5.08$, $d = 3.32$
$video_4$	Weibull with shape 0.55, scale 2700	Weibull with shape 0.18, scale 4000

The maximum likelihood estimator is $\hat{\lambda} = m/(qt_m)$. This estimator is unbiased and consistent by the strong law of large numbers ($\lim_{m\to\infty} t_m/m = 1/(q\lambda)$ a.s.).

In a similar way, the source can estimate μ if receivers probabilistically send a "goodbye" message reporting their on-time when they leave the session. Let $\tau_{m'}$ be the on-time indicated in the m'th goodbye message received at the source, then the maximum likelihood estimator of μ is simply $\hat{\mu} = m'/(\sum_{i=1}^{m'} \tau_{m'})$. The estimator $\hat{\mu}$ is unbiased and consistent.

A natural estimator for ρ is $\hat{\rho} = \mathbf{E}[\hat{N}_n]$. As long as there is no estimation of both ρ and μ, it is not possible to compute the filter coefficient A and B. Then only a naive estimator for N_n can be used, defined as the ratio of the number of ACKs received Y_n over the ACK probability p (see Section 2). Notice that $\mathbf{E}[Y_n/p] = \rho$.

We have tested the estimator $\hat{N}_n^E$ when λ and ρ are estimated. We have chosen an ACK probability $p = 0.021$, yielding $\mathbf{E}[Y_n] = 1.99$, and a hello probability $q = 0.1$, which means that, on average, one hello message is sent to the source for every 10 arrivals. The performance of the estimator can visually be observed in Fig. 3 in which five curves are plotted: (i) the original video trace, (ii) the membership estimation for the case where the parameters are known beforehand, (iii) the membership estimation for the case where estimators $\hat{\lambda} = m/(qt_m)$ and $\hat{\rho} = \mathbf{E}[Y_n]/p$ are used, (iv) the estimation returned by the EWMA algorithm (see (1)) for $\alpha = 0.99$ and (v) the estimation returned by the EWMA algorithm for $\alpha = 0.999$. Observe that when ρ and λ are estimated, the filter coefficients are computed at each observation step, whereas they are computed once for all in the other cases. As expected, when ρ and μ are unknown,

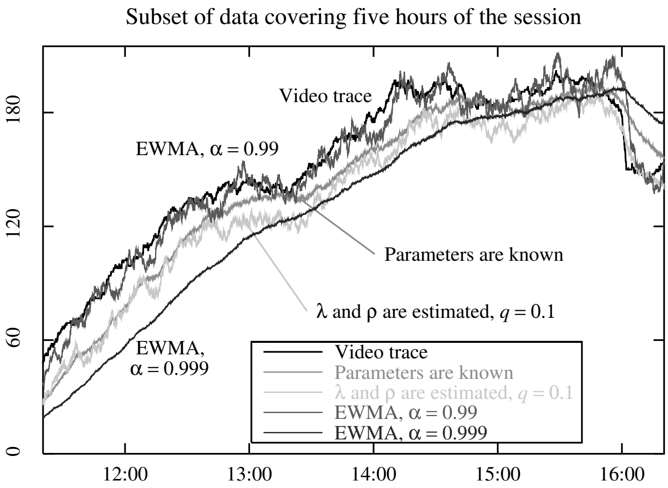

Figure 3: Membership estimation of session $video_1$ ($\rho = 94.7, p = 0.021, S = 2.5s$) when (i) parameters are known beforehand, (ii) estimators $\hat{\lambda} = m/(qt_m)$ and $\hat{\rho} = \mathbf{E}[Y_n]/p$ are used ($q = 0.1$) and (iii) EWMA estimators are used ($\alpha = 0.99, 0.999$)

Table 5: Mean and percentiles of the relative error (in %)

Estimator	Mean	25	50	75	90	95
ρ, λ known	6.0	1.2	2.6	5.0	8.8	14.5
ρ, λ estimated	5.2	1.5	3.2	5.9	10.5	16.4
EWMA $\alpha = 0.99$	4.6	1.6	3.4	6.0	9.2	11.4
EWMA $\alpha = 0.999$	6.7	1.3	3.3	7.4	14.5	21.2

Table 6: Empirical mean and variance of the estimation error

Estimator	Mean	Variance
ρ, λ known	-0.0871	26.5487
ρ, λ estimated	0.2402	37.6369
EWMA $\alpha = 0.99$	0.0006	23.1149
EWMA $\alpha = 0.999$	0.2570	79.6634

the estimator $\hat{N}_n^E$ does not behave as well as when these parameters are known beforehand. Still, its performance is reasonably fair as can be seen in Tables 5 and 6.

Table 5 reports the sample mean and some order statistics of the relative error returned by our scheme and by the EWMA algorithm proposed in (1), and Table 6 reports the sample mean and the sample variance of the error between the true membership and its estimation. Observe that, when the parameters are estimated, the relative error on $\hat{N}_n^E$ is 95% of the time within 16.4% of the true membership which is a good result (see row 3 column 7 in Table 5). As for the EWMA estimator, we observe both in Fig. 3 and Tables 5 and 6 (row 4) that the performance is very good when $\alpha = 0.99$, which is not the case when $\alpha = 0.999$ as the corresponding EWMA estimator achieves the worst performance (see row 5 in Tables 5 and 6). Notice how high is the variance of the EWMA estimator when $\alpha = 0.999$ (see row 5 column 3 in Table 6).

Remark 9.1 *For the trace $video_1$, the EWMA estimator with $\alpha = 0.99$ behaves very well in contrast to the EWMA estimator with $\alpha = 0.999$. This is exactly the inverse of what we have observed when applying both EWMA estimators on the audio trace shown in Fig. 1. There, the EWMA estimator with $\alpha = 0.99$ did not perform well, whereas the EWMA estimator with $\alpha = 0.999$ returned excellent results. In other words, given a trace, one can always find a value of α for which the EWMA estimator behaves well, but this value will be exclusive to the trace and one can not know in advance what value assign to α.*

To conclude this discussion, we believe that using the estimator $\hat{N}_n^E$ and estimating λ and ρ on-line is appealing in the sense that, even though its performance is not the best one ever, one is sure of having a fair result for a relatively small amount of ACKs. This is not the case of the EWMA estimator as not only the user will not know in advance what value assign to α, but also a "good" value for one trace is most probably not good for another.

10 Conclusion

The major contribution of this work is the design of novel estimators for evaluating the membership in multicast sessions. We have first modeled the multicast group as an $M/M/\infty$ queue and established our results under the assumption that this queue is in heavy-traffic. In this regime the backlog process of the $M/M/\infty$ queue is "close" to a diffusion process that can be used to cast our estimation problem into the appealing framework of Kalman filter theory. Using this theory we have derived an estimator that minimizes the variance of the error. Aiming at generalizing the multicast model, we relied on Wiener filter theory to compute the optimal linear estimator for session membership when the underlying model is an $M/M/\infty$ queue (the heavy traffic assumption is no longer needed). The optimality refers to the unbiasedness of the estimator and to the fact that the mean square error is minimized. The latter estimator turned out to be identical to the one designed using the Kalman filter theory. We have also developed the optimal first-order linear filter in the case where the on-time distribution is arbitrary and have derived the associated estimator in the case where the on-times have a two-stage hyperexponential distribution. The estimators have been validated on real video traces. Their performance have been shown to be excellent, one of them showing a good ability to adapt to highly dynamic multicast sessions. It is worthy to point out that it is the first time that a membership estimator is tested on real traces, exhibiting human behavior and correlations between the different processes at hand.

Acknowledgements

The authors wish to thank Profs. P. Thiran and O. Zeitouni for helpful suggestions.

References

[1] K. C. Almeroth and M. H. Ammar. *MListen*, 1995. http://www.cc.gatech.edu/computing/Telecomm/mbone/.

[2] S. Alouf. *Parameter Estimation and Performance Analysis of Several Networking Applications*. PhD thesis, University of Nice-Sophia Antipolis, November 2002. Available at http://www.inria.fr/mistral/personnel/Sara.Alouf/these.html.

[3] S. Alouf, E. Altman, C. Barakat, and P. Nain. Estimating membership in a multicast session, February 2002. INRIA, Research Report RR-4391.

[4] S. Alouf, E. Altman, and P. Nain. Optimal on-line estimation of the size of a dynamic multicast group. In *Proc. of IEEE INFOCOM '02, New York, New York*, volume 2, pages 1109–1118, June 2002.

[5] J.-C. Bolot, T. Turletti, and I. Wakeman. Scalable feedback control for multicast video distribution in the Internet. In *Proc. of ACM SIGCOMM '94, London, UK*, pages 58–67, September 1994.

[6] A. A. Borovkov. *Asymptotic Methods in Queueing Theory*. John Wiley & Sons, 1984.

[7] D. R. Cox and V. Isham. *Point Processes*. Chapman and Hall, New York, 1980.

[8] S. Deering. Host extensions for IP multicasting. RFC 1112, Network Working Group, August 1989.

[9] S. Deering. *Multicast Routing in a Datagram Internetwork*. PhD thesis, Stanford University, December 1991.

[10] C. Diot, B. N. Levine, B. Lyles, H. Kassem, and D. Balensiefen. Deployment issues for the IP multicast service and architecture. *IEEE Network magazine, Special Issue on Multicasting*, 14(1):78–88, January/February 2000.

[11] A. Dutta, H. Schulzrinne, and Y. Yemini. MarconiNet - an architecture for Internet radio and TV networks. In *Proc. of NOSSDAV '99, Basking Ridge, New Jersey*, June 1999.

[12] T. Friedman and D. Towsley. Multicast session membership size estimation. In *Proc. of IEEE INFOCOM '99, New York, New York*, volume 2, pages 965–972, March 1999.

[13] S. Haykin. *Modern Filters*. Macmillan, New York, 1989.

[14] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition, 1996.

[15] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[16] L. Kleinrock. *Queueing Systems: Theory*, volume 1. John Wiley and Sons, 1975.

[17] C. Liu and J. Nonnenmacher. Broadcast audience estimation. In *Proc. of IEEE INFOCOM '00, Tel Aviv, Israel*, volume 2, pages 952–960, March 2000.

[18] P. S. Maybeck. *Stochastic Models, Estimation and Control*, volume 1. Academic Press, New York, 1979.

[19] J. Nonnenmacher. *Reliable multicast transport to large groups*. PhD thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, July 1998.

[20] P. Robert. *Réseaux et Files d'Attente: Méthodes Probabilistes*, volume 35 of *Mathématiques & Applications*. Springer Paris, 2000.

[21] J. Rosenberg and H. Schulzrinne. Timer reconsideration for enhanced RTP scalability. In *Proc. of IEEE INFOCOM '98, San Francisco, California*, volume 1, pages 233–241, March/April 1998.

[22] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: a transport protocol for real-time applications. RFC 1889, Network Working Group, January 1996.

[23] R. F. Stengel. *Stochastic Optimal Control, Theory and Application*. John Wiley & Sons, 1986.

[24] P. Whittle. *Optimal Control. Basics and Beyond*. John Wiley & Sons, 1996.